

Regularization and decimation pseudolikelihood approaches to statistical inference in XY -spin models

Payal Tyagi^{1,2}, Alessia Marruzzo², Andrea Pagnani^{3,4}, Fabrizio Antenucci², Luca Leuzzi^{2,1}

¹ *Dipartimento di Fisica, Università Sapienza, Piazzale Aldo Moro 5, I-00185, Rome, Italy*

² *CNR-NANOTEC, Institute of Nanotechnology,*

Rome - Soft and Living Matter Lab., Piazzale Aldo Moro 5, I-00185, Rome, Italy

³ *Department of Applied Science and Technology and Center for Computational Sciences,*

Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, Italy and

⁴ *Human Genetics Foundation-Torino, Via Nizza 52, Torino, Italy*

We implement a pseudolikelihood approach with l_2 -regularization as well as the recently introduced pseudolikelihood with decimation procedure to the inverse problem in continuous spin models on arbitrary networks, with arbitrarily disordered couplings. Performances of the approaches are tested against data produced by Monte Carlo numerical simulations and compared also from previously studied fully-connected mean-field-based inference techniques. The results clearly show that the best network reconstruction is obtained through the decimation scheme, that also allows to dwell the inference down to lower temperature regimes. Possible applications to phasor models for light propagation in random media are proposed and discussed.

I. INTRODUCTION

Given a data set and a model with some unknown parameters, the inverse problem aims to find the values of the model parameters that best fit the data. In this work, in which we focus on systems of interacting elements, the inverse problem concerns the statistical inference of the underlying interaction network and of its coupling coefficients from observed data on the dynamics of the system. Versions of this problem are encountered in physics, biology (e.g., [1–3]), social sciences and finance (e.g., [4, 5]), neuroscience (e.g., [6–8]), just to cite a few, and are becoming more and more important due to the increase in the amount of data available from these fields.

A standard approach used in statistical inference is to predict the interaction couplings by maximizing the likelihood function. This technique, however, requires the evaluation of the partition function that, in the most general case, concerns a number of computations scaling exponentially with the system size. Boltzmann machine learning uses Monte Carlo sampling to compute the gradients of the Log-likelihood looking for stationary points [9] but this method is computationally manageable only for small systems. A series of faster approximations, such as naive mean-field, independent-pair approximation [7, 10], inversion of TAP equations [11, 12], small correlations expansion [13], adaptive TAP [14], adaptive cluster expansion [15] or Bethe approximations [16, 17] have, then, been developed. These techniques take as input means and correlations of observed variables and most of them assume a fully connected graph as underlying connectivity network, or expand around it by perturbative dilution. In most cases, network reconstruction turns out to be not accurate for small data sizes and/or when couplings are strong or, else, if the original interaction network is sparse.

A further method, substantially improving performances for small data, is the so-called Pseudo-Likelihood

Method (PLM) [18]. In Ref. [19] Aurell and Ekeberg performed a comparison between PLM and some of the just mentioned mean-field-based algorithms on the pairwise interacting Ising-spin ($\sigma = \pm 1$) model, showing how PLM performs sensitively better, especially on sparse graphs and in the high-coupling limit, i.e., for low temperature.

In this work, we aim at performing statistical inference on a model whose interacting variables are continuous XY spins, i.e., $\sigma \equiv (\cos \phi, \sin \phi)$ with $\phi \in [0, 2\pi)$. The developed tools can, actually, be also straightforwardly applied to the p -clock model [20] where the phase ϕ takes discretely equispaced p values in the 2π interval, $\phi_a = a2\pi/p$, with $a = 0, 1, \dots, p-1$. The p -clock model, else called vector Potts model, gives a hierarchy of discretization of the XY model as p increases. For $p = 2$, one recovers the Ising model, for $p = 4$ the Ashkin-Teller model [21], for $p = 6$ the ice-type model [22, 23] and the eight-vertex model [24–26] for $p = 8$. It turns out to be very useful also for numerical implementations of the continuous XY model. Recent analysis on the multi-body XY model has shown that for a limited number of discrete phase values ($p \sim 16, 32$) the thermodynamic critical properties of the $p \rightarrow \infty$ XY limit are promptly recovered [27, 28]. Our main motivation to study statistical inference is that these kind of models have recently turned out to be rather useful in describing the behavior of optical systems, including standard mode-locking lasers [27, 29–31] and random lasers [28, 32–35]. In particular, the inverse problem on the pairwise XY model analyzed here might be of help in recovering images from light propagated through random media.

This paper is organized as follows: in Sec. II we introduce the general model and we discuss its derivation also as a model for light transmission through random scattering media. In Sec. III we introduce the PLM with l_2 regularization and with decimation, two variants of the PLM respectively introduced in Ref. [36] and [19] for the inverse Ising problem. Here, we analyze these techniques

for continuous XY spins and we test them on thermalized data generated by Exchange Monte Carlo numerical simulations of the original model dynamics. In Sec. IV we present the results related to the PLM- l_2 . In Sec. V the results related to the PLM with decimation are reported and its performances are compared to the PLM- l_2 and to a variational mean-field method analyzed in Ref. [37]. In Sec. VI, we outline conclusive remarks and perspectives.

II. THE LEADING XY MODEL

The leading model we are considering is defined, for a system of N angular XY variables, by the Hamiltonian

$$\mathcal{H} = - \sum_{ik}^{1,N} J_{ik} \cos(\phi_i - \phi_k) \quad (1)$$

The XY model is well known in statistical mechanics, displaying important physical insights, starting from the Berezinskii-Kosterlitz-Thouless transition in two dimensions[38–40] and moving to, e.g., the transition of liquid helium to its superfluid state [41], the roughening transition of the interface of a crystal in equilibrium with its vapor [42]. In presence of disorder and frustration [43, 44] the model has been adopted to describe synchronization problems as the Kuramoto model [45] and in the theoretical modeling of Josephson junction arrays [46, 47] and arrays of coupled lasers [48]. Besides several derivations and implementations of the model in quantum and classical physics, equilibrium or out of equilibrium, ordered or fully frustrated systems, Eq. (1), in its generic form, has found applications also in other fields. A rather fascinating example being the behavior of starlings flocks [49–53]. Our interest on the XY model resides, though, in optics. Phasor and phase models with pairwise and multi-body interaction terms can, indeed, describe the behavior of electromagnetic modes in both linear and nonlinear optical systems in the analysis of problems such as light propagation and lasing [29, 54, 55]. As couplings are strongly frustrated, these models turn out to be especially useful to the study of optical properties in random media [34, 35], as in the noticeable case of random lasers [56–58] and they might as well be applied to linear scattering problems, e.g., propagation of waves in opaque systems or disordered fibers.

A. A propagating wave model

We briefly mention a derivation of the model as a proxy for the propagation of light through random linear media. Scattering of light is held responsible to obstruct our view and make objects opaque. Light rays, once that they enter the material, only exit after getting scattered multiple times within the material. In such a disordered medium, both the direction and the phase of the propagating waves are random. Transmitted light yields a

disordered interference pattern typically having low intensity, random phase and almost no resolution, called a speckle. Nevertheless, in recent years it has been realized that disorder is rather a blessing in disguise [59–61]. Several experiments have made it possible to control the behavior of light and other optical processes in a given random disordered medium, by exploiting, e.g., the tools developed for wavefront shaping to control the propagation of light and to engineer the confinement of light [62, 63].

In a linear dielectric medium, light propagation can be described through a part of the scattering matrix, the transmission matrix \mathbb{T} , linking the outgoing to the incoming fields. Consider the case in which there are N_I incoming channels and N_O outgoing ones; we can indicate with $E_k^{\text{in,out}}$ the input/output electromagnetic field phasors of channel k . In the most general case, i.e., without making any particular assumptions on the field polarizations, each light mode and its polarization polarization state can be represented by means of the 4-dimensional Stokes vector. Each t_{ki} element of \mathbb{T} , thus, is a 4×4 Müller matrix. If, on the other hand, we know that the source is polarized and the observation is made on the same polarization, one can use a scalar model and adopt Jones calculus [64–66]:

$$E_k^{\text{out}} = \sum_{i=1}^{N_I} t_{ki} E_i^{\text{in}} \quad \forall k = 1, \dots, N_O \quad (2)$$

We recall that the elements of the transmission matrix are random complex coefficients[65]. For the case of completely unpolarized modes, we can also use a scalar model similar to Eq. (2), but whose variables are the intensities of the outgoing/incoming fields, rather than the fields themselves.

In the following, for simplicity, we will consider Eq. (2) as our starting point, where E_k^{out} , E_i^{in} and t_{ki} are all complex scalars. If Eq. (2) holds for any k , we can write:

$$\int \prod_{k=1}^{N_O} dE_k^{\text{out}} \prod_{k=1}^{N_O} \delta \left(E_k^{\text{out}} - \sum_{j=1}^{N_I} t_{kj} E_j^{\text{in}} \right) = 1 \quad (3)$$

Observed data are a noisy representation of the true values of the fields. Therefore, in inference problems it is statistically more meaningful to take that noise into account in a probabilistic way, rather than looking at the precise solutions of the exact equations (whose parameters are unknown). To this aim we can introduce Gaussian distributions whose limit for zero variance are the Dirac deltas in Eq. (3). Moreover, we move to consider the ensemble of all possible solutions of Eq. (2) at given \mathbb{T} , looking at all configurations of input fields. We, thus, define the function:

$$Z \equiv \int_{\mathcal{S}_{\text{in}}} \prod_{j=1}^{N_I} dE_j^{\text{in}} \int_{\mathcal{S}_{\text{out}}} \prod_{k=1}^{N_O} dE_k^{\text{out}} \quad (4)$$

$$\times \prod_{k=1}^{N_O} \frac{1}{\sqrt{2\pi\Delta^2}} \exp \left\{ -\frac{1}{2\Delta^2} \left| E_k^{\text{out}} - \sum_{j=1}^{N_I} t_{kj} E_j^{\text{in}} \right|^2 \right\}$$

We stress that the integral of Eq. (4) is not exactly a Gaussian integral. Indeed, starting from Eq. (3), two constraints on the electromagnetic field intensities must be taken into account. The space of solutions is delimited by the total power \mathcal{P} received by system, i.e., $\mathcal{S}_{\text{in}} : \{E^{\text{in}} | \sum_k I_k^{\text{in}} = \mathcal{P}\}$, also implying a constraint on the total amount of energy that is transmitted through the medium, i. e., $\mathcal{S}_{\text{out}} : \{E^{\text{out}} | \sum_k I_k^{\text{out}} = c\mathcal{P}\}$, where the attenuation factor $c < 1$ accounts for total losses. As we will see more in details in the following, being interested in inferring the transmission matrix through the PLM, we can omit to explicitly include these terms in Eq. (6) since they do not depend on \mathbb{T} not adding any information on the gradients with respect to the elements of \mathbb{T} .

Taking the same number of incoming and outgoing channels, $N_I = N_O = N/2$, and ordering the input fields in the first $N/2$ mode indices and the output fields in the last $N/2$ indices, we can drop the “in” and “out” superscripts and formally write Z as a partition function

$$Z = \int_{\mathcal{S}} \prod_{j=1}^N dE_j \left(\frac{1}{\sqrt{2\pi\Delta^2}} \right)^{N/2} \exp \left\{ -\frac{\mathcal{H}[\{E\}; \mathbb{T}]}{2\Delta^2} \right\} \quad (5)$$

$$\mathcal{H}[\{E\}; \mathbb{T}] = - \sum_{k=1}^{N/2} \sum_{j=N/2+1}^N [E_j^* t_{jk} E_k + E_j t_{kj}^* E_k^*]$$

$$+ \sum_{j=N/2+1}^N |E_j|^2 + \sum_{k,l}^{1,N/2} E_k U_{kl} E_l^*$$

$$= - \sum_{nm}^{1,N} E_n J_{nm} E_m^* \quad (6)$$

where \mathcal{H} is a real-valued function by construction, we have introduced the effective input-input coupling matrix

$$U_{kl} \equiv \sum_{j=N/2+1}^N t_{lj}^* t_{jk} \quad (7)$$

and the whole interaction matrix reads (here $\mathbb{T} \equiv \{t_{jk}\}$)

$$\mathbb{J} \equiv \left(\begin{array}{c|c} -\mathbb{U} & \mathbb{T} \\ \hline \mathbb{T}^\dagger & -\mathbb{I} \end{array} \right) \quad (8)$$

Determining the electromagnetic complex amplitude configurations that minimize the *cost function* \mathcal{H} , Eq. (6), means to maximize the overall distribution peaked around the solutions of the transmission Eqs. (2). As the variance $\Delta^2 \rightarrow 0$, eventually, the initial set of Eqs. (2) are recovered. The \mathcal{H} function, thus, plays the role of an Hamiltonian and Δ^2 the role of a noise-inducing temperature. The exact numerical problem corresponds to the zero temperature limit of the statistical mechanical problem. Working with real data, though, which are noisy, a finite “temperature” allows for a better representation of the ensemble of solutions to the sets of equations of continuous variables.

Now, we can express every phasor in Eq. (5) as $E_k = A_k e^{i\phi_k}$. As a working hypothesis we will consider the intensities A_k^2 as either homogeneous or as *quenched* with respect to phases. The first condition occurs, for instance, to the input intensities $|E_k^{\text{in}}|$ produced by a phase-only spatial light modulator (SLM) with homogeneous illumination [67]. With *quenched* here we mean, instead, that the intensity of each mode is the same for every solution of Eq. (2) at fixed \mathbb{T} . We stress that, including intensities in the model does not preclude the inference analysis but it is out of the focus of the present work and will be considered elsewhere.

If all intensities are uniform in input and in output, this amount to a constant rescaling for each one of the four sectors of matrix \mathbb{J} in Eq. (8) that will not change the properties of the matrices. For instance, if the original transmission matrix is unitary, so it will be the rescaled one and the matrix \mathbb{U} will be diagonal. Otherwise, if intensities are *quenched*, i.e., they can be considered as constants in Eq. (2), they are inhomogeneous with respect to phases. The generic Hamiltonian element will, therefore, rescale as

$$E_n^* J_{nm} E_m = J_{nm} A_n A_m e^{i(\phi_n - \phi_m)} \rightarrow J_{nm} e^{i(\phi_n - \phi_m)}$$

and the properties of the original J_{nm} components are not conserved in the rescaled one. In particular, we have no argument, anymore, to possibly set the rescaled $U_{nm} \propto \delta_{nm}$. Eventually, we end up with the complex couplings XY model, whose real-valued Hamiltonian is written as

$$\mathcal{H} = -\frac{1}{2} \sum_{nm} J_{nm} e^{-i(\phi_n - \phi_m)} + \text{c.c.} \quad (9)$$

$$= -\frac{1}{2} \sum_{nm} [J_{nm}^R \cos(\phi_n - \phi_m) + J_{nm}^I \sin(\phi_n - \phi_m)]$$

where J_{nm}^R and J_{nm}^I are the real and imaginary parts of J_{nm} . Being \mathbb{J} Hermitian, $J_{nm}^R = J_{mn}^R$ is symmetric and $J_{nm}^I = -J_{mn}^I$ is skew-symmetric.

III. PSEUDOLIKELIHOOD MAXIMIZATION

The inverse problem consists in the reconstruction of the parameters J_{nm} of the Hamiltonian, Eq. (9).

Given a set of M data configurations of N spins $\sigma = \{\cos \phi_i^{(\mu)}, \sin \phi_i^{(\mu)}\}$, $i = 1, \dots, N$ and $\mu = 1, \dots, M$, we want to *infer* the couplings:

$$\sigma \rightarrow \mathbb{J}$$

With this purpose in mind, in the rest of this section we implement the working equations for the techniques used. In order to test our methods, we generate the input data, i.e., the configurations, by Monte-Carlo simulations of the model. The joint probability distribution of the N variables $\phi \equiv \{\phi_1, \dots, \phi_N\}$, follows the Gibbs-Boltzmann distribution:

$$P(\phi) = \frac{1}{Z} e^{-\beta \mathcal{H}(\phi)} \quad \text{where} \quad Z = \int \prod_{k=1}^N d\phi_k e^{-\beta \mathcal{H}(\phi)} \quad (10)$$

and where we denote $\beta = (2\Delta^2)^{-1}$ with respect to Eq. (4) formalism. In order to stick to usual statistical inference notation, in the following we will rescale the couplings by a factor $\beta/2$: $\beta J_{ij}/2 \rightarrow J_{ij}$. The main idea of the PLM is to work with the conditional probability distribution of one variable ϕ_i given all other variables, $\phi_{\setminus i}$:

$$\begin{aligned} P(\phi_i | \phi_{\setminus i}) &= \frac{1}{Z_i} \exp \{ H_i^x(\phi_{\setminus i}) \cos \phi_i + H_i^y(\phi_{\setminus i}) \sin \phi_i \} \\ &= \frac{e^{H_i(\phi_{\setminus i}) \cos(\phi_i - \alpha_i(\phi_{\setminus i}))}}{2\pi I_0(H_i)} \end{aligned} \quad (11)$$

where H_i^x and H_i^y are defined as

$$H_i^x(\phi_{\setminus i}) = \sum_{j(\neq i)} J_{ij}^R \cos \phi_j - \sum_{j(\neq i)} J_{ij}^I \sin \phi_j \quad (12)$$

$$H_i^y(\phi_{\setminus i}) = \sum_{j(\neq i)} J_{ij}^R \sin \phi_j + \sum_{j(\neq i)} J_{ij}^I \cos \phi_j \quad (13)$$

and $H_i = \sqrt{(H_i^x)^2 + (H_i^y)^2}$, $\alpha_i = \arctan H_i^y/H_i^x$ and we introduced the modified Bessel function of the first kind:

$$I_k(x) = \frac{1}{2\pi} \int_0^{2\pi} d\phi e^{x \cos \phi} \cos k\phi$$

Given M observation samples $\phi^{(\mu)} = \{\phi_1^{(\mu)}, \dots, \phi_N^{(\mu)}\}$, $\mu = 1, \dots, M$, the pseudo-loglikelihood for the variable i is given by the logarithm of Eq. (11),

$$\begin{aligned} L_i &= \frac{1}{M} \sum_{\mu=1}^M \ln P(\phi_i^{(\mu)} | \phi_{\setminus i}^{(\mu)}) \\ &= \frac{1}{M} \sum_{\mu=1}^M \left[H_i^{(\mu)} \cos(\phi_i^{(\mu)} - \alpha_i^{(\mu)}) - \ln 2\pi I_0(H_i^{(\mu)}) \right]. \end{aligned} \quad (14)$$

The underlying idea of PLM is that an approximation of the true parameters of the model is obtained for values that maximize the functions L_i . The specific maximiza-

tion scheme differentiates the different techniques.

A. PLM with l_2 regularization

Especially for the case of sparse graphs, it is useful to add a regularizer, which prevents the maximization routine to move towards high values of J_{ij} and h_i without converging. We will adopt an l_2 regularization so that the Pseudolikelihood function (PLF) at site i reads:

$$\mathcal{L}_i = L_i - \lambda \sum_{i \neq j} (J_{ij}^R)^2 - \lambda \sum_{i \neq j} (J_{ij}^I)^2 \quad (15)$$

with $\lambda > 0$. Note that the values of λ have to be chosen arbitrarily, but not too large, in order not to overcome L_i . The standard implementation of the PLM consists in maximizing each \mathcal{L}_i , for $i = 1 \dots N$, separately. The expected values of the couplings are then:

$$\{J_{ij}^*\}_{j \in \partial i} := \arg \max_{\{J_{ij}\}} [\mathcal{L}_i] \quad (16)$$

In this way, we obtain two estimates for the coupling J_{ij} , one from maximization of \mathcal{L}_i , $J_{ij}^{(i)}$, and another one from \mathcal{L}_j , say $J_{ij}^{(j)}$. Since the original Hamiltonian of the XY model is Hermitian, we know that the real part of the couplings is symmetric while the imaginary part is skew-symmetric. The final estimate for J_{ij} can then be obtained averaging the two results:

$$J_{ij}^{\text{inferred}} = \frac{J_{ij}^{(i)} + \bar{J}_{ij}^{(j)}}{2} \quad (17)$$

where with \bar{J} we indicate the complex conjugate. It is worth noting that the pseudolikelihood L_i , Eq. (14), is characterized by the following properties: (i) the normalization term of Eq.(11) can be computed analytically at odd with the *full* likelihood case that in general require a computational time which scales exponentially with the size of the systems; (ii) the l_2 -regularized pseudolikelihood defined in Eq.(15) is strictly concave (i.e. it has a single maximizer)[18]; (iii) it is consistent, i.e. if M samples are generated by a model $P(\phi|J^*)$ the maximizer tends to J^* for $M \rightarrow \infty$ [68]. Note also that (iii) guarantees that $|J_{ij}^{(i)} - J_{ij}^{(j)}| \rightarrow 0$ for $M \rightarrow \infty$. In Secs. IV, V we report the results obtained and we analyze the performances of the PLM having taken the configurations from Monte-Carlo simulations of models whose details are known.

B. PLM with decimation

Even though the PLM with l_2 -regularization allows to dwell the inference towards the low temperature region and in the low sampling case with better performances than mean-field methods, in some situations some cou-

plings are overestimated and not at all symmetric. Moreover, in the technique there is the bias of the l_2 regularizer. Trying to overcome these problems, Decelle and Ricci-Tersenghi introduced a new method [69], known as PLM + decimation: the algorithm maximizes the sum of the L_i ,

$$\mathcal{L} \equiv \frac{1}{N} \sum_{i=1}^N L_i \quad (18)$$

and, then, it recursively set to zero couplings which are estimated very small. We expect that as long as we are setting to zero couplings that are unnecessary to fit the data, there should be not much changing on \mathcal{L} . Keeping on with decimation, a point is reached where \mathcal{L} decreases abruptly indicating that relevant couplings are being decimated and under-fitting is taking place. Let us define by x the fraction of non-decimated couplings. To have a quantitative measure for the halt criterion of the decimation process, a tilted \mathcal{L} is defined as,

$$\mathcal{L}_t \equiv \mathcal{L} - x\mathcal{L}_{\max} - (1-x)\mathcal{L}_{\min} \quad (19)$$

where

- \mathcal{L}_{\min} is the pseudolikelihood of a model with independent variables. In the XY case: $\mathcal{L}_{\min} = -\ln 2\pi$.
- \mathcal{L}_{\max} is the pseudolikelihood in the fully-connected model and it is maximized over all the $N(N-1)/2$ possible couplings.

At the first step, when $x = 1$, \mathcal{L} takes value \mathcal{L}_{\max} and $\mathcal{L}_t = 0$. On the last step, for an empty graph, i.e., $x = 0$, \mathcal{L} takes the value \mathcal{L}_{\min} and, hence, again $\mathcal{L}_t = 0$. In the intermediate steps, during the decimation procedure, as x is decreasing from 1 to 0, one observes firstly that \mathcal{L}_t increases linearly and, then, it displays an abrupt decrease indicating that from this point on relevant couplings are being decimated[69]. In Fig. 1 we give an instance of this behavior for the 2D short-range XY model with ordered couplings. We notice that the maximum point of \mathcal{L}_t coincides with the minimum point of the reconstruction error, the latter defined as

$$\text{err}_J \equiv \sqrt{\frac{\sum_{i < j} (J_{ij}^{\text{inferred}} - J_{ij}^{\text{true}})^2}{N(N-1)/2}} \quad (20)$$

We stress that the \mathcal{L}_t maximum is obtained ignoring the underlying graph, while the err_J minimum can be evaluated once the true graph has been reconstructed.

In the next sections we will show the results obtained on the XY model analyzing the performances of the two methods and comparing them also with a mean-field method [37].

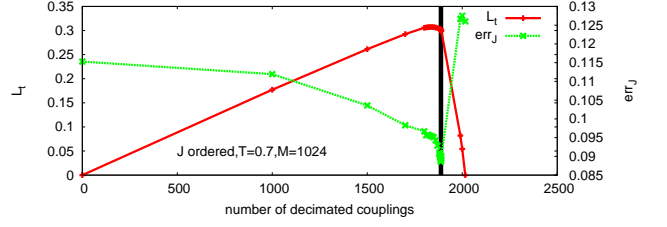


FIG. 1. The tilted likelihood \mathcal{L}_t curve and the reconstruction error vs the number of decimated couplings for an ordered, real-valued J on 2D XY model with $N = 64$ spins. The peak of \mathcal{L}_t coincides with the dip of the error.

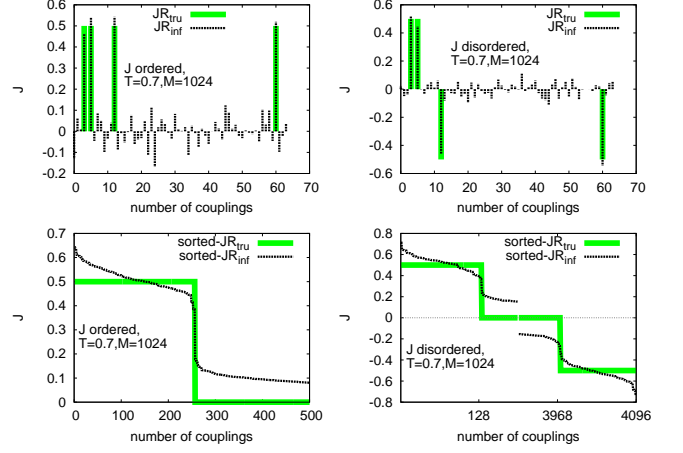


FIG. 2. Top panels: instances of single site coupling reconstruction for the case of $N = 64$ XY spins on a 2D lattice with ordered J (left column) and bimodal distributed J (right column). Bottom panels: sorted couplings.

IV. INFERRED COUPLINGS WITH PLM- l_2

A. XY model with real-valued couplings

In order to obtain the vector of couplings, J_{ij}^{inferred} the function $-\mathcal{L}_i$ is minimized through the vector of derivatives $\partial \mathcal{L}_i / \partial J_{ij}$. The process is repeated for all the couplings obtaining then a fully connected adjacency matrix. The results here presented are obtained with $\lambda = 0.01$. For the minimization we have used the MATLAB routine *minFunc_2012*[70].

To produce the data by means of numerical Monte Carlo simulations a system with $N = 64$ spin variables is considered on a deterministic 2D lattice with periodic boundary conditions. Each spin has then connectivity 4, i.e., we expect to infer an adjacency matrix with $Nc = 256$ couplings different from zero. The dynamics of the simulated model is based on the Metropolis algorithm and parallel tempering[71] is used to speed up the thermalization of the system. The thermalization is tested looking at the average energy over logarithmic time windows and the acquisition of independent configurations starts only after the system is well thermalized.

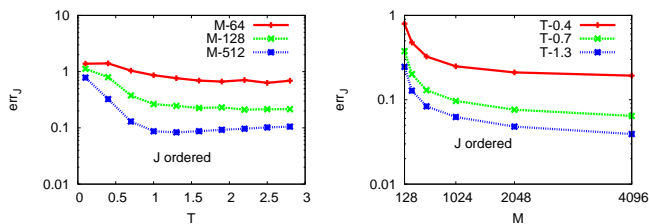


FIG. 3. Reconstruction error err_J , cf. Eq. (20), plotted as a function of temperature (left) for three values of the number of samples M and as a function M (right) for three values of temperature in the ordered system, i.e., $J_{ij} = 0, 1$. The system size is $N = 64$.

For the values of the couplings we considered two cases: an ordered case, indicated in the figure as J ordered (e.g., left column of Fig. 2) where the couplings can take values $J_{ij} = 0, J$, with $J = 1$, and a quenched disordered case, indicated in the figures as J disordered (e.g., right column of Fig. 2) where the couplings can take also negative values, i.e., $J_{ij} = 0, J, -J$, with a certain probability. The results here presented were obtained with bimodal distributed J s: $P(J_{ij} = J) = P(J_{ij} = -J) = 1/2$. The performances of the PLM have shown not to depend on $P(J)$. We recall that in Sec. III we used the temperature-rescaled notation, i.e., J_{ij} stands for J_{ij}/T .

To analyze the performances of the PLM, in Fig. 2 the inferred couplings, $\mathbb{J}_{\text{inf}}^R$, are shown on top of the original couplings, $\mathbb{J}_{\text{true}}^R$. The first figure (from top) in the left column shows the $\mathbb{J}_{\text{inf}}^R$ (black) and the $\mathbb{J}_{\text{tru}}^R$ (green) for a given spin at temperature $T/J = 0.7$ and number of samples $M = 1024$. PLM appears to reconstruct the correct couplings, though zero couplings are always given a small inferred non-zero value. In the left column of Fig. 2, both the $\mathbb{J}_{\text{inf}}^R$ and the $\mathbb{J}_{\text{tru}}^R$ are sorted in decreasing order and plotted on top of each other. We can clearly see that $\mathbb{J}_{\text{inf}}^R$ reproduces the expected step function. Even though the jump is smeared, the difference between inferred couplings corresponding to the set of non-zero couplings and to the set of zero couplings can be clearly appreciated. Similarly, the plots in the right column of Fig. 2 show the results obtained for the case with bimodal disordered couplings, for the same working temperature and number of samples. In particular, note that the algorithm infers half positive and half negative couplings, as expected.

In order to analyze the effects of the number of samples and of the temperature regimes, we plot in Fig. 3 the reconstruction error, Eq. (20), as a function of temperature for three different sample sizes $M = 64, 128$ and 512 . The error is seen to sharply rise at low temperature, incidentally, in the ordered case, for $T < T_c \sim 0.893$, which is the Kosterlitz-Thouless transition temperature of the 2XY model[72]. However, we can see that if only $M = 64$ samples are considered, err_J remains high independently on the working temperature. In the right plot of Fig. 3, err_J is plotted as a function of M for three different working temperatures $T/J = 0.4, 0.7$ and 1.3 .

As we expect, err_J decreases as M increases. This effect was observed also with mean-field inference techniques on the same model[37].

To better understand the performances of the algorithms, in Fig. 4 we show several True Positive (TP) curves obtained for various values of M at three different temperatures T . As M is large and/or temperature is not too small, we are able to reconstruct correctly all the couplings present in the system (see bottom plots). The True Positive curve displays how many times the inference method finds a true link of the original network as a function of the index of the vector of sorted absolute value of reconstructed couplings J_{ij}^{inf} . The index $n_{(ij)}$ represents the related spin couples (ij) . The TP curve is obtained as follows: first the values $|J_{ij}^{\text{inf}}|$ are sorted in descending order and the spin pairs (ij) are ordered according to the sorting position of $|J_{ij}^{\text{inf}}|$. Then, a cycle over the ordered set of pairs (ij) , indexed by $n_{(ij)}$, is performed, comparing with the original network coupling J_{ij}^{true} and verifying whether it is zero or not. The true positive curve is computed as

$$\text{TP}[n_{(ij)}] = \frac{\text{TP}[n_{(ij)} - 1] (n_{ij} - 1) + 1 - \delta_{J_{ij}^{\text{true}}, 0}}{n_{(ij)}} \quad (21)$$

As far as $J_{ij}^{\text{true}} \neq 0$, $\text{TP} = 1$. As soon as the true coupling of a given (ij) couple in the sorted list is zero, the TP curve departs from one. In our case, where the connectivity per spin of the original system is $c = 4$ and there are $N = 64$ spins, we know that we will have 256 non-zero couplings. If the inverse problem is successful, hence, we expect a steep decrease of the TP curve when $n_{ij} = 256$ is overcome.

In Fig. 4 it is shown that, almost independently of T/J , the TP score improves as M increases. Results are plotted for three different temperatures, $T = 0.4, 1$ and 2.2 , with increasing number of samples $M = 64, 128, 512$ and 1024 (clockwise). We can clearly appreciate the improvement in temperature if the size of the data-set is not very large: for small M , $T = 0.4$ performs better. When M is high enough (e.g., $M = 1024$), instead, the TP curves do not appear to be strongly influenced by the temperature.

B. XY model with complex-valued couplings

For the complex XY we have to contemporarily infer 2 apart coupling matrices, J_{ij}^R and J_{ij}^I . As before, a system of $N = 64$ spins is considered on a 2D lattice. For the couplings we have considered both ordered and bimodal disordered cases. In Fig. 5, a single row of the matrix J (top) and the whole sorted couplings (bottom) are displayed for the ordered model (same legend as in Fig. 2) for the real, J^R (left column), and the imaginary part, J^I .

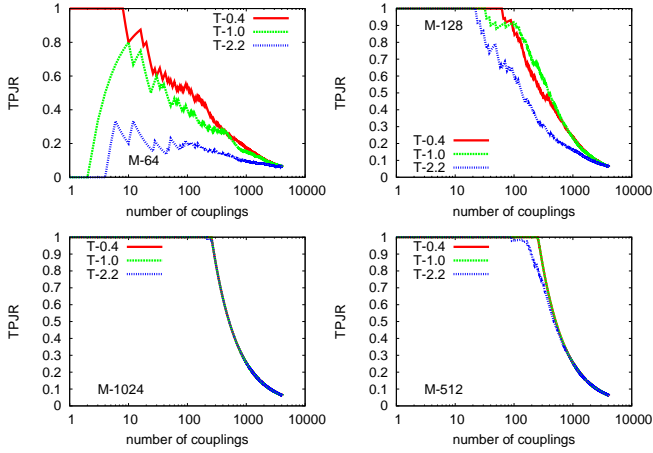


FIG. 4. TP curves for 2D short-range ordered XY model with $N = 64$ spins at three different values of T/J with increasing - clockwise from top - M .

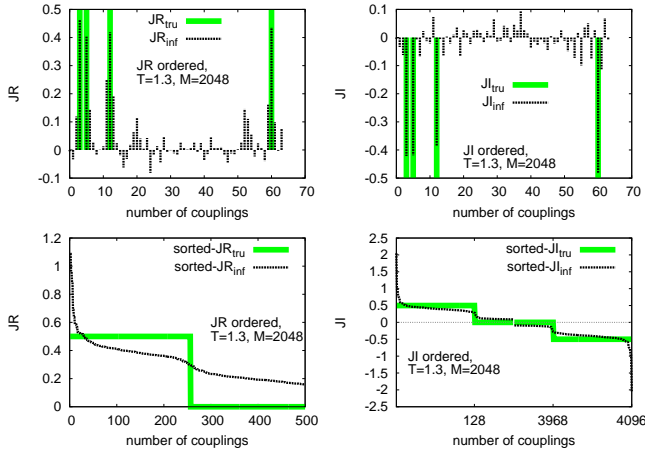


FIG. 5. Results related to the ordered complex XY model with $N = 64$ spins on a 2D lattice. Top: instances of single site reconstruction for the real, JR (left column), and the imaginary, JI (right column), part of J_{ij} . Bottom: sorted values of JR (left) and JI (right).

V. PLM WITH DECIMATION

For the ordered real-valued XY model we show in Fig. 6, top panel, the outcome on the tilted pseudolikelihood, \mathcal{L}_t Eq. (19), of the progressive decimation: from a fully connected lattice down to an empty lattice. The figure shows the behaviour of \mathcal{L}_t for three different data sizes M . A clear data size dependence of the maximum point of \mathcal{L}_t , signalling the most likely value for decimation, is shown. For small M the most likely number of couplings is overestimated and for increasing M it tends to the true value, as displayed in Fig. 7. In the bottom panel of Fig. 6 we display instead different \mathcal{L}_t curves obtained for three different values of T . Even though the values of \mathcal{L}_t decrease with increasing temperature, the value of

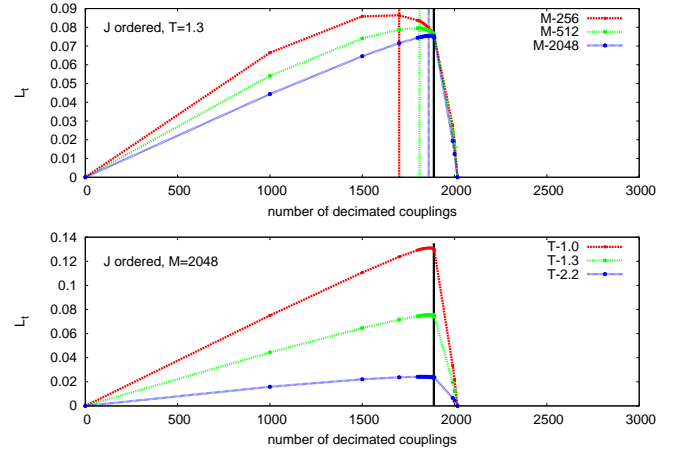


FIG. 6. Tilted Pseudolikelihood, \mathcal{L}_t , plotted as a function of decimated couplings. Top: Different \mathcal{L}_t curves obtained for different values of M plotted on top of each other. Here $T = 1.3$. The black line indicates the expected number of decimated couplings, $x^* = (N(N-1) - Nc)/2 = 1888$. As we can see, as M increases, the maximum point of \mathcal{L}_t approaches x^* . Bottom: Different \mathcal{L}_t curves obtained for different values of T with $M = 2048$. We can see that, with this value of M , no differences can be appreciated on the maximum points of the different \mathcal{L}_t curves.

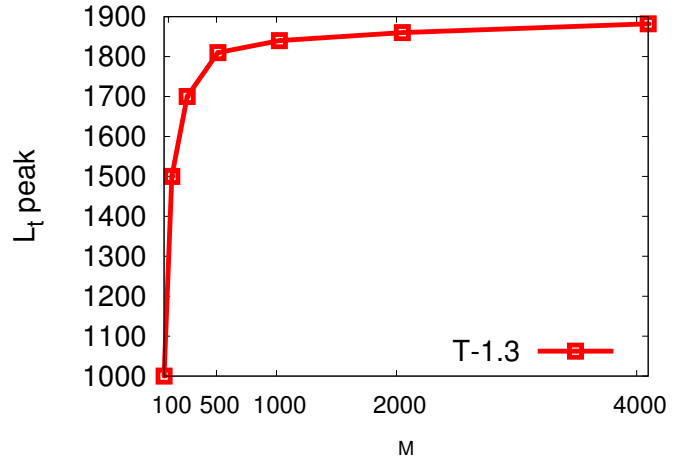


FIG. 7. Number of most likely decimated couplings, estimated by the maximum point of \mathcal{L}_t , as a function of the number of samples M . We can clearly see that the maximum point of \mathcal{L}_t tends toward x^* , which is the right expected number of zero couplings in the system.

the most likely number of decimated couplings appears to be quite independent on T with $M = 2048$ number of samples. In Fig. 8 we eventually display the tilted pseudolikelihood for a 2D network with complex valued ordered couplings, where the decimation of the real and imaginary coupling matrices proceeds in parallel, that is, when a real coupling is small enough to be decimated its imaginary part is also decimated, and vice versa. One can see that though the apart errors for the real and imagi-

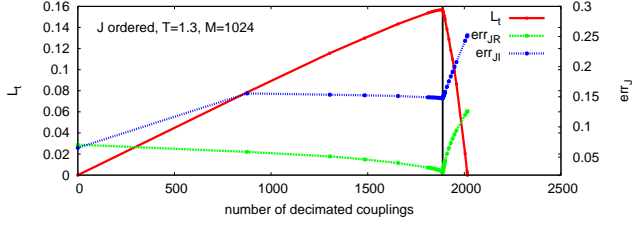


FIG. 8. Tilted Pseudolikelihood, \mathcal{L}_t , plotted with the reconstruction errors for the XY model with $N = 64$ spins on a 2D lattice. These results refer to the case of ordered and complex valued couplings. The full (red) line indicates \mathcal{L}_t . The dashed (green) and the dotted (blue) lines show the reconstruction errors (Eq. (20)) obtained for the real and the imaginary couplings respectively. We can see that both err_{JR} and err_{JI} have a minimum at x^* .

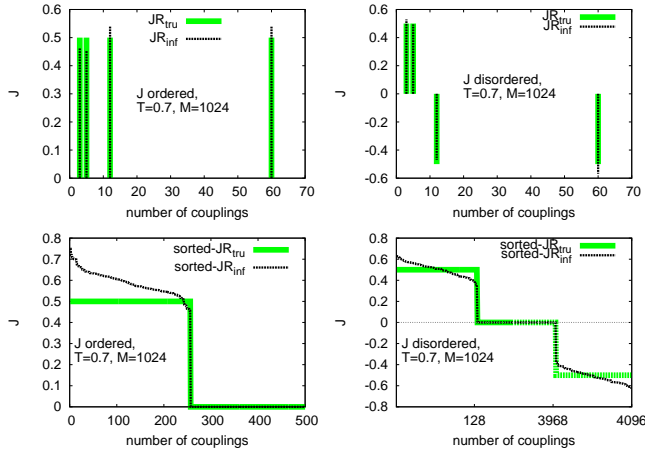


FIG. 9. XY model on a 2D lattice with $N = 64$ sites and real valued couplings. The graphs show the inferred (dashed black lines) and true couplings (full green lines) plotted on top of each other. The left and right columns refer to the cases of ordered and bimodal disordered couplings, respectively. Top figures: single site reconstruction, i.e., one row of the matrix J . Bottom figures: couplings are plotted sorted in descending order.

nary parts are different in absolute values, they display the same dip, to be compared with the maximum point of \mathcal{L}_t .

Once the most likely network has been identified through the decimation procedure, we perform the same analysis displayed in Fig. 9 for ordered and then quenched disordered real-valued couplings and in Fig. 10 for complex-valued ordered couplings. In comparison to the results shown in Sec. IV, the PLM with decimation leads to rather cleaner results. In Figs. 12 and 11 we compare the performances of the PLM with decimation in respect to ones of the PLM with l_2 -regularization. These two techniques are also analysed in respect to a mean-field technique previously implemented on the same XY systems[37].

For what concerns the network of connecting links, in

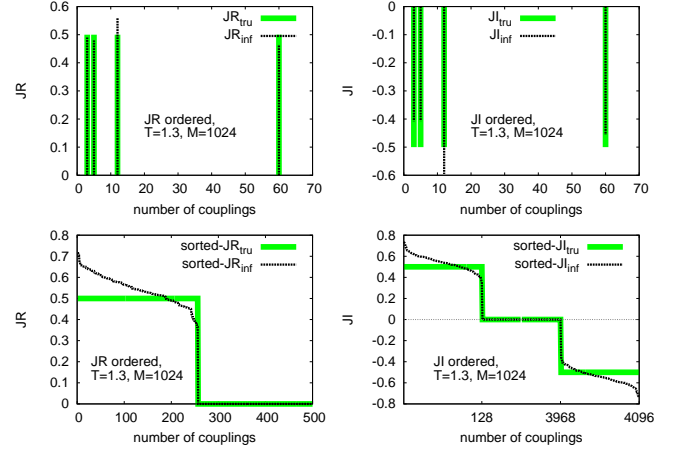


FIG. 10. XY model on a 2D lattice with $N = 64$ sites and ordered complex-valued couplings. The inferred and true couplings are plotted on top of each other. The left and right columns show the real and imaginary parts, respectively, of the couplings. Top figures refer to a single site reconstruction, i.e., one row of the matrix J . Bottom figures report the couplings sorted in descending order.

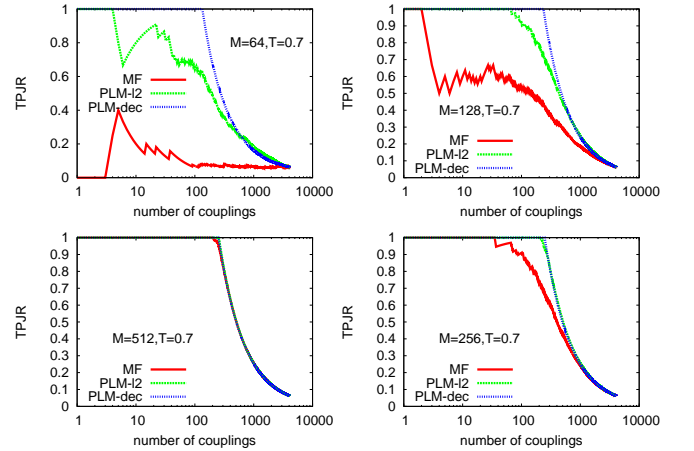


FIG. 11. True Positive curves obtained with the three techniques: PLM with decimation, (blue) dotted line, PLM with l_2 regularization, (green) dashed line, and mean-field, (red) full line. These results refer to real valued ordered couplings with $N = 64$ spins on a 2D lattice. The temperature is here $T = 0.7$ while the four graphs refer to different sample sizes: M increases clockwise.

Fig. 11 we compare the TP curves obtained with the three techniques. The results refer to the case of ordered and real valued couplings, but similar behaviours were obtained for the other cases analysed. The four graphs are related to different sample sizes, with M increasing clockwise. When M is high enough, all techniques reproduce the true network. However, for lower values of M the performances of the PLM with l_2 regularization and with decimation drastically overcome those ones of the previous mean field technique. In particular, for

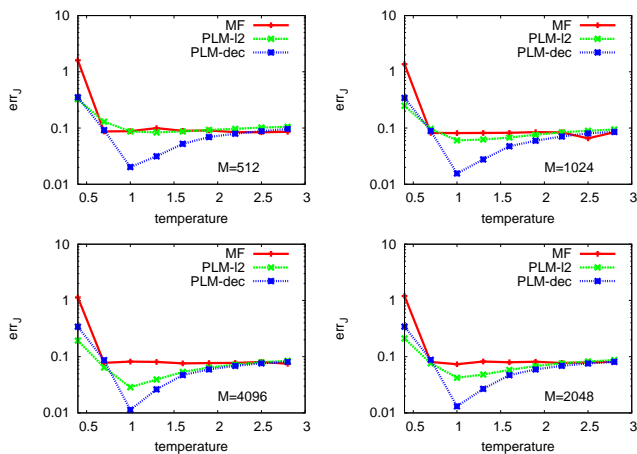


FIG. 12. Variation of reconstruction error, err_J , with respect to temperature as obtained with the three different techniques, see Fig. 11, for four different sample size: clockwise from top $M = 512, 1024, 2048$ and 4096 .

$M = 256$ the PLM techniques still reproduce the original network while the mean-field method fails to find more than half of the couplings. When $M = 128$, the network is clearly reconstructed only through the PLM with decimation while the PLM with l_2 regularization underestimates the couplings. Furthermore, we notice that the PLM method with decimation is able to clearly infer the network of interaction even when $M = N$ signalling that it could be considered also in the under-sampling regime $M < N$. In Fig. 12 we compare the temperature behaviour of the reconstruction error. It can be observed that for all temperatures and for all sample sizes the reconstruction error, err_J , (plotted here in log-scale) obtained with the PLM+decimation is always smaller than that one obtained with the other techniques. The temperature behaviour of err_J agrees with the one already observed for Ising spins in [73] and for XY spins in [37] with a mean-field approach: err_J displays a minimum around $T \simeq 1$ and then it increases for very lower T ; however, the error obtained with the PLM with decimation is several times smaller than the error estimated by the other methods.

VI. CONCLUSIONS

Different statistical inference methods have been applied to the inverse problem of the XY model. After

a short review of techniques based on pseudo-likelihood and their formal generalization to the model we have tested their performances against data generated by means of Monte Carlo numerical simulations of known instances with diluted, sparse, interactions.

The main outcome is that the best performances are obtained by means of the pseudo-likelihood method combined with decimation. Putting to zero (i.e., decimating) very weak bonds, this technique turns out to be very precise for problems whose real underlying interaction network is sparse, i.e., the number of couplings per variable does not scale with number of variables. The PLM + decimation method is compared to the PLM + regularization method, with l_2 regularization and to a mean-field-based method. The behavior of the quality of the network reconstruction is analyzed by looking at the overall sorted couplings and at the single site couplings, comparing them with the real network, and at the true positive curves in all three approaches. In the PLM + decimation method, moreover, the identification of the number of decimated bonds at which the tilted pseudo-likelihood is maximum allows for a precise estimate of the total number of bonds. Concerning this technique, it is also shown that the network with the most likely number of bonds is also the one of least reconstruction error, where not only the prediction of the presence of a bond is estimated but also its value.

The behavior of the inference quality in temperature and in the size of data samples is also investigated, basically confirming the low T behavior hinted by Nguyen and Berg [73] for the Ising model. In temperature, in particular, the reconstruction error curve displays a minimum at a low temperature, close to the critical point in those cases in which a critical behavior occurs, and a sharp increase as temperature goes to zero. The decimation method, once again, appears to enhance this minimum of the reconstruction error of almost an order of magnitude with respect to other methods.

The techniques displayed and the results obtained in this work can be of use in any of the many systems whose theoretical representation is given by Eq. (1) or Eq. (9), some of which are recalled in Sec. II. In particular, a possible application can be the field of light waves propagation through random media and the corresponding problem of the reconstruction of an object seen through an opaque medium or a disordered optical fiber [59–63, 65–67].

-
- [1] S. Balakrishnan, H. Kamisetty, J. G. Carbonell, S.-I. Lee, and C. J. Langmead, Proteins: Struct., Funct., and Bioinf. **79**, 1061 (2011).
 - [2] M. Ekeberg, C. Lövkqvist, Y. Lan, M. Weigt, and E. Aurell, Phys. Rev. E **87**, 012707 (2013).

- [3] C. Feinauer, M. J. Skwark, A. Pagnani, and E. Aurell, PLoS Comput Biol **10**, 1 (2014).
- [4] I. Mastromatteo, E. Zarinelli, and M. Marsili, JSTAT, P03011 (2012).
- [5] S. Yamanaka, M. Ohzeki, and A. Decelle,

- Journal of the Physical Society of Japan **84**, 024801 (2015), <http://dx.doi.org/10.7566/JPSJ.84.024801>.
- [6] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, *Nature* **440**, 1007 (2006).
 - [7] Y. Roudi, J. Hertz, and E. Aurell, *Front. Comp. Neuros.* **3**, 22 (2009).
 - [8] J. Tyrcha, Y. Roudi, M. Marsili, and J. Hertz, *JSTAT*, P03005 (2013).
 - [9] K. P. Murphy, *Machine Learning. A Probabilistic Perspective* (2012).
 - [10] Y. Roudi, J. Tyrcha, and J. Hertz, *Phys. Rev. E* **79**, 051915 (2009).
 - [11] H. J. Kappen and F. B. Rodríguez, *Neural Comp.* **10**, 1137 (1998).
 - [12] T. Tanaka, *Phys. Rev. E* **58**, 2302 (1998).
 - [13] V. Sessak and R. Monasson, *J. Phys. A* **42**, 055001 (2009).
 - [14] *Advanced mean field methods: theory and practice*, Neural information processing series, Saad, David and Oppen, Manfred (MIT press, Breckenridge, CO US, 2001).
 - [15] S. Cocco and R. Monasson, *Journal of Statistical Physics* **147**, 252 (2012).
 - [16] F. Ricci-Tersenghi, *JSTAT* **2012**, P08015 (2012).
 - [17] H. C. Nguyen and J. Berg, *JSTAT* **2012**, P03004 (2012).
 - [18] P. Ravikumar, M. J. Wainwright, and J. D. Lafferty, *Ann. Statist.* **38**, 1287 (2010).
 - [19] E. Aurell and M. Ekeberg, *Phys. Rev. Lett.* **108**, 090201 (2012).
 - [20] R. Potts, *Proc. Camb. Phil. Soc.* **48**, 106 (1952).
 - [21] J. Ashkin and E. Teller, *Phys. Rev.* **64**, 178 (1943).
 - [22] L. Pauling, *J. Am. Chem. Soc.* **57**, 2680 (1935).
 - [23] R. J. Baxter, *Exactly solved models in statistical mechanics* (Academic Press Inc. (London), 1982).
 - [24] B. Sutherland, *J. Math. Phys.* **11**, 3183 (1970).
 - [25] C. Fan and F. Y. Wu, *Phys. Rev. B* **2**, 723 (1970).
 - [26] R. Baxter, *Phys. Rev. Lett.* **26**, 832 (1971).
 - [27] A. Marruzzo and L. Leuzzi, *Phys. Rev. B* **91**, 054201 (2015).
 - [28] A. Marruzzo and L. Leuzzi, arXiv:1512.02153 and *Phys. Rev. B*, in production (2016).
 - [29] A. Gordon and B. Fischer, *Phys. Rev. Lett.* **89**, 103901 (2002).
 - [30] O. Gat, A. Gordon, and B. Fischer, *Phys. Rev. E* **70**, 046108 (2004).
 - [31] L. Angelani, C. Conti, L. Prignano, G. Ruocco, and F. Zamponi, *Phys. Rev. B* **76**, 064202 (2007).
 - [32] L. Angelani, C. Conti, G. Ruocco, and F. Zamponi, *Phys. Rev. Lett.* **96**, 065702 (2006).
 - [33] L. Leuzzi, C. Conti, V. Folli, L. Angelani, and G. Ruocco, *Phys. Rev. Lett.* **102**, 083901 (2009).
 - [34] F. Antenucci, C. Conti, A. Crisanti, and L. Leuzzi, *Phys. Rev. Lett.* **114**, 043901 (2015).
 - [35] F. Antenucci, A. Crisanti, and L. Leuzzi, *Phys. Rev. A* **91**, 053816 (2015).
 - [36] B. Schölkopf, J. C. Platt, and T. Hoffman, eds., *Proceedings from the conference, "Neural Information Processing Systems 2006*, Advances in Neural Information Processing Systems (2006).
 - [37] P. Tyagi, A. Pagnani, F. Antenucci, M. Ibañez Berganza, and L. Leuzzi, *JSTAT* **P05**, 031 (2015).
 - [38] V. L. Berezinskii, *Zh. Eksp. Teor. Fiz.* **59**, 907 (1970).
 - [39] V. L. Berezinskii, *Sov. Phys. JETP* **32**, 493 (1971).
 - [40] J. Kosterlitz and D. Thouless, *J.Phys.C* **5**, L124 (1972).
 - [41] E. Brézin, *J. Phys. (France)* **43**, 15 (1982).
 - [42] J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, Cambridge, 1996).
 - [43] J. Villain, *J. Phys. C* **10**, 1717 and 4793 (1977).
 - [44] E. Fradkin, B. A. Huberman, and S. H. Shenker, *Phys. Rev. B* **18**, 4789 (1978).
 - [45] Y. Kuramoto, *Lect. N. Phys.* **39**, 420 (1975).
 - [46] S. Teitel and C. Jayaprakash, *Phys. Rev. B* **27**, 598 (1983).
 - [47] S. Teitel and C. Jayaprakash, *Phys. Rev. Lett.* **51**, 1999 (1983).
 - [48] M. Nixon, E. Ronen, A. A. Friesem, and N. Davidson, *Phys. Rev. Lett.* **110**, 184102 (2013).
 - [49] C. Reynolds, *Computer Graphics* **21**, 25 (1987).
 - [50] J. L. Deneubourg and S. Goss, *Ethology, Ecology, Evolution* **1**, 295 (1989).
 - [51] A. Huth and C. Wissel, in *Biological Motion*, edited by W. Alt and E. Hoffmann (Springer-Verlag, Berlin, 1990) pp. 577–590.
 - [52] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, *Phys. Rev. Lett.* **75**, 1226 (1995).
 - [53] A. Cavagna, S. M. D. Queirós, I. Giardina, F. Stefanini, and M. Viale, *Proceedings of the Royal Society of London B: Biological Sciences* **280** (2013).
 - [54] F. Antenucci, M. Ibañez Berganza, and L. Leuzzi, *Phys. Rev. A* **91**, 043811 (2015).
 - [55] F. Antenucci, M. Ibañez Berganza, and L. Leuzzi, *Phys. Rev. B* **92**, 014204 (2015).
 - [56] D. S. Wiersma, *Nature Physics* **4**, 359 (2008).
 - [57] J. Andreasen, A. A. Asatryan, L. C. Botten, B. A. Byrne, H. Cao, L. Ge, L. Labonté, P. Sebbah, A. D. Stone, H. E. Türeci, and C. Vanneste, *Adv. Optics and Photonics* **3**, 88 (2011).
 - [58] F. Antenucci, A. Crisanti, M. Ibañez Berganza, A. Marruzzo, and L. Leuzzi, arXiv:1509.06955 and *Phil. Mag.*, in production (2016).
 - [59] I. Vellekoop and A. Mosk, *Opt. Lett.* **32**, 2309 (2007).
 - [60] I. Vellekoop, E. van Putten, A. Lagendijk, and A. Mosk, *Opt. Express* **16**, 67 (2008).
 - [61] I. M. Vellekoop and A. P. Mosk, *Phys. Rev. Lett.* **101**, 120601 (2008).
 - [62] H. Yilmaz, W. L. Vos, and A. P. Mosk, *Biomed Opt Express* **4**, 1759 (2013).
 - [63] R. F., N. Caselli, S. Vignolini, and *et al.*, *Nature Materials* **13**, 720 (2014).
 - [64] J. W. Goodman, *Statistical Optics* (John Wiley & Sons, 1985).
 - [65] S. M. Popoff, G. Leroosey, R. Carminati, M. Fink, A. C. Boccara, and S. Gigan, *Phys. Rev. Lett.* **104**, 100601 (2010).
 - [66] D. Akbulut, T. J. Huisman, E. G. van Putten, W. L. Vos, and A. P. Mosk, *Opt. Expr.* **19**, 4017 (2011).
 - [67] S. M. Popoff, G. Lerosy, M. Fink, A. C. Boccara, and S. Gigan, *New J. Phys.* **13**, 123021 (2011).
 - [68] J. Besag, *J. Royal Stat. Soc. D* **24**, 179 (1975).
 - [69] A. Decelle and F. Ricci-Tersenghi, *Phys. Rev. Lett.* **112**, 070603 (2014).
 - [70] M. Schmidt, “minfunc: unconstrained differentiable multivariate opti (2005).
 - [71] D. J. Earl and M. W. Deem, *Phys. Chem. Chem. Phys.* **7**, 3910 (2005).
 - [72] P. Olsson and P. Minnhagen, *Phys. Scr.* **43**, 203 (1992).
 - [73] H. C. Nguyen and J. Berg, *Phys. Rev. Lett.* **109**, 050602 (2012).